

THE MEDICAL AND SOCIETAL CONSEQUENCES OF THE HUMAN GENOME PROJECT

Dr. Francis Collins

[Note: Dr. Collins' speech is inaudible and unclear in some parts due to technical difficulties with the recording of his presentation.]

DR. COLLINS: Thank you, Jan, for a very fine introduction. I'm sure my parents would have been delighted to hear all those nice things you said about me, although they might have disagreed with a few of them. It is wonderful to have a chance to be here with you all this beautiful Saturday. And, I am enormously impressed with the number of folks that have turned out for NIH to tell you something about what we do here.

As I was coming here, I saw the traffic jam out there on Rockville Pike. And, initially, I thought, my gosh, there must be tens of thousands of people. Well, some of them were going to the household recycling pick-up over at the parking lot over [inaudible, laughter]. And, maybe after they recycle they'll come on over and fill this room even further.

But this is wonderful to see all of you here today. I do feel somewhat badly for the people way back where who I can sort of vaguely see the outlines of. And, I hope you can hear my voice. But, I suspect that when I start putting visual images up here on the screen, they are going to be a bit hard for you to make out. I will try my best in this presentation to describe what's up there on the screen in words so that it won't be absolutely essential to follow what I have to say. But, I'd also say, if there are those way in the back who want to try to get a little closer, you could probably rearrange things, bring your chair forward, whatever you'd like to do, I won't be offended if [inaudible].

Well, you've already had a full morning, a lot of presentations about areas of research that NIH is vigorously involved in. I'm here to tell you about one in particular, which has been particularly exciting over the course of the past few months. Actually, it's been building over the course of about 10 years. And, what I want to do is to tell you briefly what the human genome project is and where we are with this grand adventure, then what it is going to do for the practice of medicine, which is the reason we are doing it in the first place.

I also want to address some of the ethical, legal, and social issues that arise as a consequence of our ability to read our own instruction book, which is what we are about in this project. I hope to talk for no more than about 30 minutes and I hope we can then have some questions and answers at the end, because I'd be very interested in some of your thoughts and dreams about this as well. And, I [inaudible] this whole thing is more fun for me, as well. So, that's what we are going to do. Again, I'll thank you Jan, for organizing today and inviting me. I'm delighted to have a chance to be part of this program.

Well, here's your best chance to figure out whether you can see the slides or not. Probably many of you are going, oh, uh-oh. Well, don't worry, I'll describe what's there. This is a cartoon from the *Cincinnati Inquirer* from about the first of the year as we went into the new millennium. And, what you see here is a canoe riding along the [inaudible]... labeled the human genome project. We are comparing the genome project with Lewis and Clarke. We are exploring unknown territory. We are going off into places that we have not previously had the ability to look at. And I hope you will share that sense here, of adventure; it's an adventure we are all on.

I started to venture out into the Northwest territory, or some other new geographic area; maybe it's a bit like going out into space and searching to see what we could do out there. But, in fact, this is different because this is an adventure into ourselves. What we are doing here is reading our blueprint, taking a look at those instructions, and specifying all of the biological properties of the human being. We only have to do this once and we are in the midst of this right now. Some people working with [inaudible] everything, it will change. As a result, much of it already has. And, we'll never go back again to the phase where we didn't have the information. It is going to speed things up a lot.

Why are we doing this project? Well, pure and simple, I'm a physician, and the reason I'm interested in this effort is because I believe—and many of the folks that I work with share this belief—that uncovering the mysteries of our genetic instruction book is going to lead us to answers for virtually every disease, because virtually every disease has at least some hereditary upholding.

Now think about what happens, what occurs in your family. Almost all of us would think of some disorder that seems to occur in our families at a somewhat higher frequency than it does in other folks, at least as we perceive it. And, there is a good chance that those particularly [inaudible] risk somewhere in the DNA sequences of you and your relatives.

Now, people have tended, in the past, I think, to think of genetics as the study of relatively rare conditions, the very [inaudible], very straightforward way. Things like Huntington's disease, or cystic fibrosis. But genetics is much broader than that. The major causes of disease and sickness and premature death and, a lot of times, civilization, diabetes, cancer, heart disease, all have genetic underpinnings. But, [inaudible] clear about this. They are not part of why [inaudible] DNA. It's not that you are doomed to a certain health [inaudible]. But, there are previous conditions that [inaudible]. And those interact with each other and with the environment and the choices that you make about lifestyle and [inaudible]. And, all together, all of those things may result in your getting or not getting a disease like diabetes.

Genetics has this powerful attribute to it: you don't have to know what the answer

is going to be when you start. And, so, many folks trying to understand diabetes or cancer or heart disease are taking the genetics approach as a way of shining the light into an area that's previously not been well understood.

Diabetes is a good example. My own research lab over there in Building 9 has a team of young scientists trying to understand the hereditary factors of diabetes. You might say to me, "Well wait a minute, we already understand diabetes, don't we?" Well, yes and no. We understand that something's awry within certain glucose metabolism. We understand some of the predisposing factors. We know that heredity is certainly a compelling part of it because [inaudible] other families. But, not only the clean-cut dominant or recessive fashion that [inaudible] would have approved of. It's a very messy kind of an [inaudible].

We don't understand, really, when you look closely at the fundamental issues, why diabetes comes about. What's wrong? Why is [inaudible]? Why is resistance to insulin apparently part of the problem, as well? If we could identify the specific genes that are involved in diabetes predisposition, we have the best clue ever for why this would be [inaudible].

Just three weeks ago, the first one of those genes was identified by this strategy that I'm going to tell you about. This sets the precedent for unraveling the mysteries of diseases like diabetes, as well as cancer, the major mental illnesses, multiple sclerosis, and asthma; all of those are conditions that have significant genetic contributions. Even though, let me say it again, those are predispositions and not necessarily predeterminations.

Even infectious diseases like AIDS have some genetic contributions; the exposure to the virus is one thing, but what you happen to do with that virus once you are exposed is another. There is probably, in this room, about three people who, if they were exposed to HIV virus over and over again, would never get sick, because they are genetically immune to this disease that has [inaudible] that the virus needs to get inside the cell and wreak its havoc.

An understanding of how that works is one of the major focuses right now: trying to develop an approach to make everybody resistant to AIDS, to figure out how that 1% of the population does it. Maybe we could offer that to everybody else.

But genetics has something to offer for virtually every disease. My one exception might be cases of [inaudible], when [inaudible] didn't have a lot to do with it. But, I'll tell you, in some cases of [inaudible] exist.

Now, if you want to understand genetics, this molecule called DNA is the thing that you need to understand. And, again, I apologize for the people in the back, you

probably can't see DNA right now, but it is this wonderful molecule [inaudible] figured out [inaudible] years ago. And it conveys information in a very simple and in a very elegant way. Science is very satisfying in this regard. For me, somebody who is trained originally in physics and chemistry and thought biology was sort of [inaudible], it was wonderful to discover that it isn't, the DNA, as a way in which [inaudible]. It is an enormously precise, digital, intellectually satisfying kind of way of [inaudible].

So, DNA works basically by storing information, and it stores information in a simple alphabet with only four letters: A, C, G and T. So your entire [inaudible] is written in this rather unusual code of four letters, in a particular order, of A, C, G and T. But if you go [inaudible], I would ask you to guess how many of those letters it would take to specify a human being. Because those letters have to be positioned [inaudible] to direct everything that has to happen from the time you were just a fertilized single cell embryo to where you are right now.

That's a lot of information to carry around. It seems almost unimaginable that there could be a [inaudible] set of information. But it is a [inaudible] set. And maybe some people are guessing various numbers; I won't ask you to stand up and shout them out, but I'll tell you the answer. The answer is three billion. Three billion. Three billion of those letters is the human genetic code, the human genome, as we call it, the genome that feeds all of the DNA. By the way, an [inaudible] is there, a genome [inaudible] just about the same size, three billion letters. Obviously, organized in some different way.

There is a tendency for simple organisms to have smaller genomes, but it is not followed very precisely. The world record holder in terms of genome size, weighing in at 100 billion letters of the DNA code, which may surprise you, is [inaudible]. [Inaudible] a beautiful and complicated, but, it doesn't seem like this justifies [inaudible]. But, I'm sorry, that's the data. I [inaudible] is full, of course, a [inaudible] that we call [inaudible] maybe they don't really need all those hundred billions to do all the things that they do.

There are a lot of extra [inaudible] floating around. Well, [inaudible] and I'm sorry to say there are a lot floating around our [inaudible]. It seems, from what we know right now, that as much as 90% of our three billion letters may not be doing very much, just coming along for the ride. Well, you could argue, and you'd be right, but, we don't know enough to say that that 90% is [inaudible]. It may well be and we are just not smart enough [inaudible].

But we want to understand this molecule. Now how does DNA work? DNA sits inside the nucleus of the cell. It is the instruction book, but somebody has to carry out the instructions, right? That is done by this process of RNA [inaudible]. The way this works in a simple cartoon is, you can think of DNA as an instruction packet for a grocery cart. The instructions are carried out by RNA. So you see [inaudible] out of the DNA grocery cart here in the [inaudible]. [Inaudible] carried by RNA out to the place where [inaudible]. That's what these do, they [inaudible] and you end up with a protein that's going to use [inaudible] important function [inaudible]. This is a flow of information. DNA is an instruction book, [inaudible] is carrying

out the instructions.

So how does a genetic disease come about? That's the basis of this kind of a [inaudible]. Well, it comes about because there is a misspelling of the DNA. Now it is not A, C, G, T, but it is C, C, G, T, which means there will be a misspelling in the RNA, which is just a copy of that. And then, when you go to the protein [inaudible] it may not turn out quite as well. In this situation you end up with a protein that has a glitch, and it lacks some of the oomph that it is supposed to have to carry out its functions. So, I would ask you, am I now talking about an occasional person who has this kind of a glitch? Well, I'm sorry to tell you if you came here today thinking you were the perfect genetic specimen, I have bad news for you. Because there aren't any.

At the DNA level, we are all afflicted with a variety of [inaudible] that have been passed down to us by previous generations. There are no perfect specimens. The estimates are that we are all walking around with somewhere between five and [inaudible] spelling changes, somewhere in those three billion letters, that places us at risk for something. Now many of those things you'll never find out about. You might have [inaudible] because of your family history. You won't find out about many of them because the [inaudible] you might need a combination of such particular glitches before you are really at significant risk. Or, you might be in an environment where you happen to manage to avoid a trigger of some sort.

So, these flaws are not necessarily immediately apparent by your medical history or your family history but some of them may be for the [inaudible] they need to cause you trouble. So, I'll ask you at this point in the presentation to think for a minute about what you would like as far as access to that information. If, at the end of the lunch here, I set up a booth out there in the hall, and said, oh, come on [inaudible], and I'll scrape your cheek with a wooden [inaudible] and get a few cells from you cheek that has your DNA in it, and I'll call you up next week and tell you what your glitches are, would you want to know?

If I can give you that kind of a printout saying here is a list of things that you are at higher-than-average risk for, I could probably also tell you here is a list of things you are at lower-than-average risk for. Would you want to know? How many people would want to know the information? [Inaudible] not bad. How many would not want to know? [Inaudible].

I suspect that you are trying to answer that. You are thinking to yourself, "Now wait a minute, you are kind of pulling a fast one on me here, you haven't told me what I'm going to do with this information. Is it going to help me? Are you going to say, you are at risk for Alzheimer's disease, have a nice day?" [Laughter] "Or, are you going to say, you are at risk for Alzheimer's disease, and here is what you can do to reduce that risk by following this particular diet or lifestyle or medical surveillance." Then there would probably be a lot more interest in it. I didn't tell you that part.

And, I think, in general, as we've asked this question in focus groups and gone through a lot of discussion over the last few years, most people's interest in the information is closely tied to whether there is something they can do with it in order to improve their health. That makes a lot of sense. [inaudible]. There are situations though, and I'll come back to these, where people

already can get this information, can do something useful with it to reduce their risk, and are afraid to get the information because of concerns that it might be used against them. This is a serious issue so I'll come back this.

So, okay, you are interested in knowing your flaws. How are we going to find this out? Well, the human genome project basically was designed for medical purposes, to understand those three billion letters, to identify the genes involved in virtually every disease, and to give us the ability to make precise predictions about risks and to develop new treatments that we never could have thought of without understanding the disease at this kind of level.

The project got underway in 1990. The initial goal was to try to have the sequence of the human genome, those three billion letters, completed by 2005. And as you heard in the introduction, we are going to beat that goal by a significant degree. And, I'm happy to say, not only are we ahead of schedule, we are [inaudible].

[Laughter]. [Inaudible]. [Applause]. We are really about 25% less than the original cost prediction of all this that we put forward 10 years ago.

We started out, not trying to actually sequence every letter, because we needed to have technology that would be good enough for that. Instead, [inaudible] sort of a view from 30,000 feet because you start counting the [inaudible] the [inaudible] developed maps of this human chromosome over the course of the first five years.

About a year and a half ago we looked to see where we were with the mapping and the sequencing effort and these were the human chromosomes. There are 24 of them and the areas that are colored in were the areas that we had sequences for in March of 1999. By that time, we had gotten to the point where we had confidence in our technology and confidence in our investigators, who were some of the world's brightest scientists.

Sixteen centers have been working on this effort. I've had the honor of serving as their private manager. We decided that we would pull out all the stops and see what we could do over the course of, say, a year and a half. By June 2000 -- this is what it looks like -- basically every chromosome hovered around the 90% level. So we got to go to the White House on June 26th and made a big announcement with President Clinton standing there cheering us on. We had a live satellite link to the United Kingdom, where Prime Minister Blair was also making remarks about the significance of this moment, because Britain has also been a big part of it, as have France, Germany, Japan, and China.

It was a pretty remarkable milestone to contemplate, that on June 26, 2000, we basically said, "We've got most of the information, 90% of it." And all the information derived by this public [inaudible] around the country and around the world, has been put up on the Internet every 24 hours. We decided back in 1996 that this information was so valuable, so useful, that we shouldn't have it [inaudible] away, anybody [inaudible] was a good idea to try and use it as soon as this was derived.

So that's been our plan, that every 24 hours the new data goes up on the Internet. If you have Internet access, you can go look at a lot of A, C, G and T, about three billion of them, and see if you could help us sort out what they mean.

We also made a decision early on that we would not have any of the centers involved in this effort by unpacking on the genes that they discovered, because we felt that this was potentially going to be a deterrent for the use of the information. And so all of it is placed in a public domain [inaudible] by this mechanism of putting it up on the Internet which then makes it officially public domain.

I have to say, there has been much interest from the private sector in doing this kind of DNA sequencing as well. We welcome it. [Inaudible] different model, where private businesses aimed to derive a lot of [inaudible] information and by the necessities of the business world, [inaudible]. We understand that. And I think there are many places in which the biotechnology of the pharmaceutical industry is going to be absolutely crucial for the future of this deal, if we are going to see the kind of products that the public wants to come out of it.

So, while I have this [inaudible] between the private and public sectors, I can tell you quite honestly that's been a great deal of [inaudible]. We are actually working quite well together in this particular project. [Inaudible] in producing things in each [inaudible] that I think everybody is happy to see happen.

This cartoon here is a couple of [inaudible] standing and staring at a very large pile of [inaudible]. And one of them is staring at the box that says [inaudible] genome. And the other one says, "I think I found a corner piece." [Laughter]. This is my way of saying, having those three billion letters is not the end of the story. [Inaudible] is the end of the beginning, this whole [inaudible]. We have now crossed the threshold of not knowing this information to knowing it. But we still have to figure out how to read this language. And that's going to take a great deal of effort and enterprise over the course of the next many years, involving investigators all over the world, putting their best ideas into place here to try to understand what this book is telling us.

There are a variety of ways that we are going to try to do that. [Inaudible] study the genome of other model organisms, because the comparisons are extremely useful in understanding how genes work. We already have complete sequences of the fruit fly, the roundworm, a bacteria -- many bacteria, actually -- [inaudible], and we'll soon have a sequence of the [inaudible], which is a very useful comparison [inaudible]. If you look at genes and [inaudible], look at our sequence, and look at that on [inaudible] 90% the same.

If you look at the part of the genome that doesn't seem to be doing very much, you won't see much similarity at all. It's a very useful way to pick out the parts of the genome that are probably most important to study. People are proposing sequencing lots of other genomes. There is a lobby out there, right now, trying to say the next genome could be the cure for muscular dystrophy, the [inaudible] is only 1.6% different than [inaudible], remarkably similar to the sequence of 98.4% identity. It's just enough. Obviously, that 1.6% is pretty interesting. We'd like to know what it is and how it works. There's a lot more sequencing going on even though the human part of this is pretty much done. It will take another couple of years to get the

human sequencing [inaudible] accurate form [inaudible] graph, a highly useful graph, but it is still a graph. And we don't want to leave it as a graph; it is our instruction book. We want to get it right for all [inaudible].

There are a number of other issues on this [inaudible]. I won't go into it, but I hope you will appreciate that the genome project is not over, because it has essentially just begun. We have the first major database of explanations, the sequence of the genome. Now our big [inaudible] is to figure out what it means.

Now, what it means is a way of heading toward the medical advances that are the most important aspects of this. Let me tell you, in general, what the extent of their task is going to be as a consequence of the genome project for your medical field.

We will, in the course of the next five to six years, [inaudible] uncover those genetic contributions that play a role in the hereditary predisposition to virtually all kinds of disorders. [Inaudible] like to do that. Each discovery of a genetic change will give us a chance to make predictions about who is at risk, so this hypothetical booth that I was going to set up out here in the hallway won't be so hypothetical in six or seven years.

But, of course, as we discussed a few minutes ago, that kind of information is mostly valuable if it allows you to practice preventive medicine. The preventive medicine is designed for use. That's really what this would look like, individualized. What we currently do is sort of a one size fits all. We tell everybody, well, you have to have this piece of [inaudible] at this age and you have to follow this particular diet and go through this kind of exercise program and medical surveillance. And we are doing that, sort of averaging the whole population together. But, in some instances, that isn't right for a particular person.

If you are kind of [inaudible] cancer because you have a glitch somewhere in an important gene that's involved in [inaudible], you should not be waiting until you are age 50 to get your first [inaudible], you should probably be starting that and colonoscopy at age 35 or 40. In so doing, we might pick up that [inaudible] and get it out of there before its already become a cancer and it's too late.

This is the [inaudible] of the, sort of, third phase of genomic metaphase which you use this predictive information to allow people to reduce their risks and practice good surveillance of their lifestyle or dietary [inaudible]. But that's actually sort of a birth phase. In another 10 or 15 years, what you will see is the array of treatment that we have for the genes will have changed very dramatically. Because we'll take advantage of these incredible insights that come out of genetic studies to design a whole new cohort of drugs that are much more precisely targeted to the primary problem, as opposed to some downstream system. Therefore, it is more likely to be successful and less likely to have side effects.

We will also learn how to prescribe drugs, not in such a generic way where a person with a particular diagnosis always gets the same drug, but in a way that is individualized. Not everybody with a high blood draw probably needs exactly the same intervention. If we understood and collected all the data, we'd be able to pick the intervention more precisely for that person. So, it's probably not unimaginable, but in 10 or 15 years, before the [inaudible]

prescription, well, first one of [inaudible] for genome and see if that's the right specific for you [inaudible].

Let me give you a [inaudible] of what the future might look like [inaudible] general practitioner. Now, [inaudible] your family here. Walter and Ruth are dad and mom, Helen is their daughter, and they are coming to see their practitioner in the year 2010 because they all have recently developed asthma, and naturally you've had a pretty tough time with it. There is no family history of asthma; it sort of came along [inaudible]. Well, the doctors are working up a situation where you could have these treatments. There are a variety of things that you might choose for [inaudible]. Furthermore, Walter and Ruth, the parents, have not previously had a lot of opportunities for thinking about their medical [inaudible]. [Inaudible].

But they are at a moment now when you started [inaudible] perhaps possible to fight. But this, the general practitioner, might give them a [inaudible]. The GP says, "You know, it's 2010; that Dr. Collins, back in the cafeteria at Natcher 10 years ago, said there might be an opportunity here for people to find out their risks. Well, we know he was right, and we have the tests available that will give you all a chance to find out what your future risks are. Do you want to know?" And Walter and Ruth say, "Yes, I guess so if there is something we can do about it. But, maybe, you don't need to tell us about those conditions where you don't have anything to offer."

Meanwhile, of course, we have to pay attention to Helen, [inaudible] treat it, but there is also in 2010 a [inaudible] to make a prediction about which of the asthma drugs that are available might be best suited for her. Well, that's [inaudible] something for Helen, [inaudible] Walter and Ruth have been tested for a dozen conditions and for interventions that are now available to cut down the [inaudible] risks. Well, with Helen, it turns out, the genetic test indicates that the usual inhalant, which is the first [inaudible] treatment for asthma in both children and adults, [inaudible] is probably not going to work for her.

Now, I didn't make this up. This is based on a study that was published about a month ago, and it is rather clear that there is a subset of individuals who don't respond very well to this inhaler and you can [inaudible] because they have a variance in a particular gene that's involved in the way that the drug works. So it may be better for Helen to be treated in a different way. This is not at all science fiction. I don't [inaudible] a different therapy is more successful to get her asthma quickly under control.

Walter and Ruth get the information as well. For Walter, it turns out that he may [inaudible] cancer, even though he doesn't have a [inaudible]. So he might make the decision that he will begin colonoscopy when he's 40, which is a very effective way, if you knew the people were high risk, of prevention.

Ruth turns out to be infected with a disease that I bet you never heard of, [inaudible] common and [inaudible]. You probably don't know you have it, but it's a gene called [inaudible] protein. The genes were [inaudible] overly vigilant, and you continue to [inaudible] iron into your body even after your iron stores are [inaudible]. And the iron deposits in the heart and the liver and the pancreas [inaudible] and ultimately causes significant problems, particularly with

[inaudible].

So this is a disease that [inaudible] you could [inaudible] but [inaudible] cause trouble. If the gene has been found, they say it's not relatively straightforward to identify the roughly more than 300 people who, like Ruth, have this condition, but are totally unaware of it. And here it is with [inaudible] situation, in terms of making an argument for the value of the information, and the actual treatment can not be simpler.

The treatment is that you go to the Red Cross and you give a unit of blood a little more often than anybody else. And that takes enough iron out of your system to keep your balance okay. And that's not only effective, it's free, and you are allowed to do something that makes you feel good at the same time, by giving something that somebody else can use.

So it is very likely that in the course of the next few years, you will find yourself being offered an opportunity, [inaudible]. Ruth, in my hypothetical situation, turns out to be one of those people affected, [inaudible] and probably within a great deal of difficulty that she would have had in her 50's, 60's and 70's with heart disease or liver trouble.

But this is not a far-fetched scenario. I tend to be a little optimistic [inaudible] that the course of this kind of research is really so [inaudible] that it is, I think, more likely that I'm being pessimistic here than I'm being optimistic about the timetable.

I will say, however, that the one cloud above all of this that might prevent this kind of scenario from happening is the general sense that many people have of being worried about [inaudible] of genetic information. I mentioned that earlier and now at the end here I want to come back to that. The genome project, from its outset, has had a component of, focus on the ethical, legal, and social implications of this research. I don't know what we propose can be brand new dilemmas by doing this project, but we greatly accelerated a pace of which this approach is happening and therefore the number of people whose lives are going to have a direct effect at this type of accelerated information.

We have identified a number of very significant [inaudible] and some of them are further along than others [inaudible]. [Inaudible] about this notion of getting genetic information: [inaudible] the number one issue that people point to is how is that going to be viewed?, who is going to know about it?, and what is it going to do to me if they find out I'm at risk for something? And that is a serious issue. Could you lose your health insurance because you are found to be at risk for something based on the genetic test? So, likely, [inaudible]

[END OF TAPE]